AD-A104 182 CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF STATISTICS F/G 12/2
TRANSFORMING CONTINGENCY TABLES.(U)
AUG 81 M M MEYER N00014-80-C-0637
UNCLASSIFIED TR-214 NL

1 OF 1
AD A
104182

END
DATE
FILMED
10-81
DTIC

LEVEL II

AD A104182

DTIC ELECTE SEP 14 1981 S D H

DEPARTMENT

OF

STATISTICS

1975

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DTIC FILE COPY

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

81 9 14 031

TRANSFORMING CONTINGENCY TABLES

by

Michael M. Meyer

Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213

Technical Report No. 214

August, 1981

1. Introduction

There are many numerical procedures for calculating the maximum likelihood estimates for loglinear models of frequency data. The most popular methods are the Iterative Proportional Fitting Procedure (IPFP) and variants of Newton's method. For problems involving a large number of parameters Newton's method is often impractical. On the other hand many models can not be expressed in a form which allows the simple IPFP to be applied. In these circumstances some other nonlinear optimization technique (e.g. the Generalized Iterative Scaling method of Darroch and Ratcliff (1972) or the extensions of the IPFP due to Haberman (1974)) must be used. As the basic IPFP is a well understood, robust, and widely available algorithm it would often be desirable to cajole a given problem into a form where the IPFP can be applied. We present a general theorem on transforming contingency tables and several applications where the transformation technique has allowed us to take advantage of the IPFP and resulted in simple and useful procedures. A further advantage of this technique is that it is sometimes possible to recognize closed-form estimates in the transformed problem while they would be overlooked in the original setting.

We shall view the estimation problem as one of minimizing the Kullback-Leibler information distance between two probability mass functions (p.m.f.'s) and will roughly follow the notation of Csiszár (1976). Although we have adopted the information distance point of view, the duality between maximum likelihood estimation and minimum information estimation (see e.g. Darroch and Ratcliff (1972)) implies that the results of this paper can just as well be interpreted from the maximum likelihood point of view.

2. Background and Notation

Csiszár (1976) presents a very elegant discussion of the IPFP by developing a "geometry" for the information measure. A simplified version of the chief results of this theory are outlined below. Let n, p, q, r, s, and t denote p.m.f.'s which are non-zero for all elements of a finite set I. The Kullback-Leibler information number (or directed divergence) specifies a distance,

$$I(p\|q) = \frac{1}{|I|}\sum_{i\varepsilon I} p(i)\ \ln\ (p(i)/q(i))$$

between p and q. The principle of minimum discriminant information, as formulated by Kullback (1959), aims to minimize the distance between a reference distribution, q above, and a family of other distributions. The properties of such estimates have been studied extensively. The most important results can be found in Kullback (1959) and are summarized, with a special emphasis on contingency tables, in Gokhale and Kullback (1978).

We next develop an appropriate family, $\mathcal{E}$, of p.m.f.'s. A convex set, $\mathcal{E}$, of p.m.f.'s is called linear if when p and q are in $\mathcal{E}$ and $t = \alpha \cdot p + (1-\alpha) \cdot q$ ($\alpha \varepsilon \mathbb{R}$) is a p.m.f., then t is also in $\mathcal{E}$. A p.m.f. which satisfies

$$I(q\|r) = \min_{p\varepsilon \mathcal{E}} I\ (p\|r)$$

is called the I-projection of r on $\mathcal{E}$ and will be denoted by $q = \mathbb{P}_{\mathcal{E}}(r)$. Csiszár gives conditions under which $\mathbb{P}_{\mathcal{E}}(r)$ exists (it is always unique) and develops a geometry for I-projections by using an analogue of Pythagorous' Theorem. Now let $\mathcal{F} = \{f_\gamma : \gamma\varepsilon\Gamma\}$ be a set of real valued functions on I and $\mathcal{A} = \{a_\gamma : \gamma\varepsilon\Gamma\}$ be real constants. Define $\mathcal{M}_{\mathcal{F}}$ to be span ($\mathcal{F}$). A linear set, $\mathcal{E}$, can be constructed by considering the set of p for which,

$$\sum_{i\varepsilon I} p(i) \cdot f_\gamma(i) = a_\gamma \quad ; \quad \gamma\varepsilon\Gamma \quad .$$

When we consider s to be an observed probability function and

$$a_\gamma = \sum_{i\in I} s(i)\, f_\gamma(i) \quad : \quad \gamma\in\Gamma$$

then the duality between maximum likelihood and minimum discriminant estimation states that if

$$\hat{q} = \mathbb{P}_{E}(r)$$

then

$$\ln(\hat{q}) \in M_F + \ln(r)$$

and

$$\hat{q} - s \in M_F^{\perp} \quad ,$$

i.e. $\hat{q}$ is the m.l.e. (under Poisson sampling) for the corresponding log-affine model. Csiszár's principle theorem says that if E is the finite intersection of the linear sets E_k (i.e. $E = \bigcap_{k\in K} E_k$) then $\hat{q} = \mathbb{P}_E(r)$ is the pointwise limit of $q_n = \mathbb{P}_{E_n}(q_n)$ $n = 1,2,3$ where $q_o = r$ and $E_n = E_i$ if $i = n \mod |K|$.

Example 1. Ordered Categories

Let p be an observed 3x3 probability function obtained via multinomial sampling and consider the ordered categories model

$$\mathbb{E}(p_{ij}) = q_{ij} \quad ,$$

and $$\ln(q_{ij}) = \alpha_i + \beta_j + j\cdot\gamma_i + i\cdot\delta_j \quad ; i,j = 1,2,3 \quad .$$

The linear manifold for this model is spanned by a set of tables, f_R^i, f_{OR}^i, f_C^j and f_{OC}^j; $i,j = 1,2,3$. The subscripts R, OR, C and OC indicate that the vector corresponds to Row, Ordered Row, Column or Ordered Column parts of the model, while the superscript indicates the row or column number, e.g.,

$$f_R^2 = \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 1 & 1 & 1 \\ \hline 0 & 0 & 0 \\ \hline\end{array} \qquad f_{OR}^2 = \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 0 & 1 & 2 \\ \hline 0 & 0 & 0 \\ \hline\end{array}$$

$$f_C^1 = \begin{array}{|c|c|c|}\hline 1 & 0 & 0 \\ \hline 1 & 0 & 0 \\ \hline 1 & 0 & 0 \\ \hline\end{array} \qquad f_{OC}^3 = \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \hline 0 & 0 & 2 \\ \hline\end{array}$$

The general structure is that f_R^i (or f_C^j) is a table of zeros except for the i'th row (j'th column) which contains ones, i.e.,

$$f_R^i(k, \ell) = \begin{cases} 1 & k = i \\ 0 & k \neq i \end{cases} .$$

Similarly, for the ordered row and column tables, the general form is

$$f_{OC}^j(k, \ell) = \begin{cases} k-1 & \ell = j \\ 0 & \ell \neq j \end{cases} .$$

We now group the spanning tables into sets of related constraints. Let

$$F_R = \{f_R^i, f_{OR}^i : i = 1,2,3\}$$

and

$$F_C = \{f_C^j, f_{OC}^j : j = 1,2,3\} .$$

The sets of constants, A_R and A_C, are determined by the inner products of p with the spanning vectors.

The linear spaces of p.m.f.'s corresponding to these constraints and constants are:

$$\mathcal{E}_R = \{\text{p.m.f.'s } p \text{ s.t. } \sum_{k,\ell} f_A^i(k,\ell)\cdot p(k,\ell) = a_A^i \quad ; \quad A = R,OR;\ i = 1,2,3\}$$

$$\mathcal{E}_C = \{\text{p.m.f.'s } p \text{ s.t. } \sum_{k,\ell} f_B^j(k,\ell)\cdot p(k,\ell) = a_B^j \quad ; \quad B = C,OC;\ j = 1,2,3\}$$

In order to find the M.L.E.'s of cell probabilities for this model we need to be able to compute $\hat{q} = \mathbb{P}_{\mathcal{E}}(r)$ for $r(k,\ell) = 1, \forall k, \ell$ and $\mathcal{E} = \mathcal{E}_R \cap \mathcal{E}_C$. The theory tells us that this I-projection can be obtained by cyclically projecting onto $\mathcal{E}_R$ and $\mathcal{E}_C$.

3. Motivation for Transformations

As algorithms for the basic IPFP are widely available, it is often advantageous for us to be able to pose a problem in a way that makes it amenable to attack by means of these programs.

A very simple example, which is prototypical of those that will arise in our later discussion, can be constructed as follows.

Example 2

Consider a triple of observed counts $z = (z_1, z_2, z_3)$ from 3 independent Poisson random variables with mean $m = (m_1, m_2, m_3)$ and having observed values (1, 3, 5). Suppose we wish to fit the log-affine model,

$$\ln(m) \in \ln \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + \mathcal{M}$$

where $$\mathcal{M} = \operatorname{span} \left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\} .$$

It is a simple matter to verify that the M.L.E. is $\hat{m} \doteq (.694, 3.611, 4.694)$. Now consider the related contingency table

$$z^* = \begin{array}{|c|c|} \hline 2z_1 & z_2 \\ \hline z_2 & 2z_3 \\ \hline \end{array} = \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 3 & 10 \\ \hline \end{array}$$

and the model for the mean, m^* ,

$$\ln(m^*) \in \mathcal{M}^*$$

where $\mathcal{M}^* = \text{span}\left\{\begin{pmatrix}1\\1\\0\\0\end{pmatrix}\begin{pmatrix}0\\0\\1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\\0\end{pmatrix}\begin{pmatrix}0\\1\\0\\1\end{pmatrix}\right\}$, the "independence"

manifold. This model has a closed-form M.L.E., namely,

$$\hat{m}^* = \begin{array}{|c|c|}\hline 5\times5/18 & 5\times13/18 \\ \hline 5\times13/18 & 13\times13/18 \\ \hline\end{array} \doteq \begin{array}{|c|c|}\hline 1.389 & 3.611 \\ \hline 3.611 & 9.389 \\ \hline\end{array}$$

Now note that

$$\hat{m}^* = \begin{array}{|c|c|}\hline 2\hat{m}_1 & \hat{m}_2 \\ \hline \hat{m}_2 & 2\hat{m}_3 \\ \hline\end{array} .$$

In other words it is possible to fit the "difficult" model, $\mathcal{M}$, by transforming the table and fitting the "easy" model, $\mathcal{M}^*$, to the transformed table. In the process of doing this transformation we have also recognized that the original log-affine model actually had closed-form estimates, namely

$$\hat{m}_1 = (2z_1 + z_2)^2 \,/\, (4 \times (z_1 + z_2 + z_3)$$

$$\hat{m}_2 = (2z_1 + z_2)(2z_3 + z_2) \,/\, (4 \times (z_1 + z_2 + z_3))$$

$$\hat{m}_3 = (2z_3 + z_2)^2 \,/\, (4 \times (z_1 + z_2 + z_3)) .$$

This example is clearly contrived to please Dr. Pangloss. We shall later present a more realistic version with similar consequences. ■

In the preceding example we transformed the data into a form where it was much easier to compute the M.L.E. of the vector of expected values. Of course we have yet to prove that the above manipulation is any more than a numerical coincidence; such proofs are the subject of this paper.

The idea of modifying a problem so that it is amenable to analysis by existing or easier methods is not at all new. An old example of this phenomenon is the method of filling in missing values to transform an "unbalanced" analysis of variance into a "balanced" problem. Although fitting an ANOVA model to an incomplete data array is conceptually easy, the calculations are much simpler when the missing values are filled in. The same is true of Example 2. Fitting the model M is not difficult but the model M^* is much simpler.

For such a small problem as Example 2 there is little practical advantage to be gained from the transformation technique. The motivation for this research lies in some very large problems considered by Fienberg and Wasserman (1981). We discuss their examples and some related theory in section 5.

Thus far we have not given any motivation for the data transformation of Example 2. We now continue the example and give a heuristic justification of the method and at the same time present a more realistic version of this problem.

Example 2 (continued)

Let us consider a general log-affine model for the Poisson data, z, with mean value, m, namely

$$\ln(m) \in \ln(d) + \mathcal{M}$$

where d is any fixed triple of positive numbers and $\mathcal{M}$ is as before. Note that if d is the vector of all ones then this reduces to a simple log-linear model. Regardless of d , a version of the sufficient statistics for this model are

$$v_1 = 2z_1 + z_2$$

and

$$v_2 = z_2 + 2z_3 .$$

Now consider the table z^* as a transformation, g , of z , i.e.

g maps $Z^3 \to Z^{2\times 2}$ such that

$$\begin{pmatrix} z^*_{11} \\ z^*_{12} \\ z^*_{21} \\ z^*_{22} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

We now note that $z^*_{1+} = z^*_{+1} = v_1$ and $z^*_{2+} = z^*_{+2} = v_2$. In other words the sufficient statistics for the data z with model $\mathcal{M}$ are represented twice in the margins of z*. Thus if we fit the row and column margins model, $\mathcal{M}^*$, to z* we might expect the the likelihood equation for model $\mathcal{M}$ is also satisfied. This turns out to be the case, but we have ignored the question of whether $\hat{m}$ satisfies the log-affine model. We shall see that if we fit the log-affine model

$$\ln(m^*) \in \ln(g(d)) + \mathcal{M}^* \tag{3.1.1}$$

to the data z* then the M.L.E., $\hat{m}$, can be recovered. The simple IPFP, with starting table g(d), will converge to the M.L.E. $\blacksquare$

In section 4 we discuss what conditions are necessary to justify procedures such as those discussed above.

4. A Transformation Theorem

We present a collection of conditions (graniloquently labelled as a theorem) relating to how one may transform estimation problems. First we consider a very weak condition which will be used in the theorem and which is itself sometimes useful.

The idea of this first result is that it is often possible to fortuitously solve a difficult estimation problem by "accidentally" satisfying the conditions. Consider the problem

$$\text{maximize} \quad f(m|z)$$

$$\text{subject to} \quad m \in \mathcal{D}$$

where $\mathcal{D}$ is some constraint space. Assume f has a unique maximum over $\mathcal{D}$

and denote the maximizing m by $\tilde{m}$. Now consider the problem

$$\text{maximize} \quad f(m|z)$$
$$\text{subject to} \quad m \in \mathcal{D}^\dagger$$

where $\mathcal{D}^\dagger \supset \mathcal{D}$. Denote the maximizing m by $\tilde{m}^\dagger$. It is a trivial observation that <u>if</u> $\tilde{m}^\dagger \in \mathcal{D}$ then $\tilde{m}^\dagger = \tilde{m}$. In other words, if the maximizing value, $\tilde{m}^\dagger$, under the weaker conditions, $\mathcal{D}^\dagger$, happens to satisfy the stronger conditions, $\mathcal{D}$, then $\tilde{m}^\dagger$ is also the maximizer under the stronger conditions. Notice also that we did not require $\tilde{m}^\dagger$ to be unique as the uniqueness of $\tilde{m}$ implies there is at most one $\tilde{m}^\dagger$ in $\mathcal{D}$. This idea could be used anywhere a constrained maximum is required but there is no <u>guarantee</u> that $\tilde{m}^\dagger$ <u>will</u> be in $\mathcal{D}$. We will use this general idea in frequency data circumstances where we can <u>prove</u> that $\tilde{m}^\dagger$ will be in $\mathcal{D}$ and where the constraints $\mathcal{D}^\dagger$ are easier to deal with than the constraints $\mathcal{D}$.

We now turn to a more refined version of this method. The statement of the result is in terms of the Kullback-Leibler distance but could equally be stated in terms of the (dual) likelihood function.

<u>Theorem</u>

Let g be a one to one mapping of the p.m.f.'s on a set I into the p.m.f.'s on a set I^*. If E is a linear set of p.m.f.'s on I, then define $g(E) = \{g(p): p \in E\}$. Let E^* be a linear set of p.m.f.'s on I^* such that $g(E) \subset E^*$. If g is such that

$$(4.1) \qquad I(p||q) = k \cdot I(g(p)||g(q)) \text{ for } p,q \in E ,$$

and if $\qquad \mathbb{P}_{E^*}(g(r)) \in g(E), \qquad$ then

$$\mathbb{P}_E(r) = g^{-1}(\mathbb{P}_{E^*}(g(r))) \qquad \blacksquare$$

The condition (4.1) could be generalized to allow $I(p||q) = f(I(g(p)||g(q)))$ where f is any monotone one to one mapping. We have no need for such generality here.

The theorem shows that under certain conditions it is possible to calculate an I-projection in a transformed table and then invert the transformation to obtain the I-projection in the original setting. Verifying the conditions of the theorem may itself be a difficult task. There are at least two ways of using the theorem. In some situations it may be possible to define the linear set E^* so that $g(E) = E^*$. This is the easier case and it essentially just relabels the problem. However even such simple relabeling can be helpful in interpreting the model or recognizing, say, a model in the transformed space for which closed form estimates are known to exist. The second application of the theorem requires more work to verify the conditions, but is also more generally applicable. Here we take a linear set E^* which is much larger than $g(E)$, but we then need to prove that $P_{E^*}(g(r)) \in g(E)$. In other words, even though E^* contains $g(E)$ we need to show that for any $g(r)$, the I-projection onto E^* is always an element of $g(E)$. For a particular set of data it may be easy to verify this condition. All we need do is fit the transformed model and see if the I-projection is in $g(E)$. To prove this type of result for a general class of problems is more difficult. We will illustrate the simple case of the theorem with the following examples. Section 5 will be devoted to a discussion of a set of examples where $g(E) \subset E^*$.

<u>Example 3</u>

This example is a continuation of Example 1. The problem concerns a 3 x 3 table where the classifying variables have a natural ordering. The specific model we consider fits row and column margins and linearly-weighted row and column margins.

We have previously shown that the row and column constraints can be considered in pairs and each of the pairs of constraints can be individually fit. Thus if (w_1, w_2, w_3) are the current fitted values for, say, the first row, we need to adjust this triple so that its row and ordered row margins match some specified constants.

Let E_S be the set of positive triples which satisfy the row and ordered row constraints for the first row, i.e.,

$$E_S = \Big\{\text{positive triples, } q : 2q_1 + q_2 = 2a_R^1 - a_{OR}^1 \stackrel{D}{=} a_3 \quad \text{and} \quad q_2 + 2q_3 = a_{OR}^1 \stackrel{D}{=} a_4\Big\}$$

Now consider the function

$$g : w \to \begin{array}{|c|c|} \hline w_1 & \frac{1}{2}\, w_2 \\ \hline \frac{1}{2}\, w_2 & w_3 \\ \hline \end{array}$$

and define

$$\begin{aligned} E^* &= g(E_S) \\ &= \Big\{2 \times 2 \text{ tables } \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array} \text{ such that } a + b = a + c = \tfrac{1}{2}\, a_3 \text{ and } \\ &\qquad d + c = d + b = \tfrac{1}{2}\, a_4\Big\} . \end{aligned}$$

Note that the constraints on $\mathcal{E}^*$ imply that b equals c which means that g^{-1} is well defined on $\mathcal{E}^*$. It is not a difficult calculation to verify that $I(q||w) = I(g(q)||g(w))$. Our theorem now allows us to calculate $\mathbb{P}_{\mathcal{E}_S}(w)$ as $g^{-1}\,\mathbb{P}_{\mathcal{E}^*}(g(w))$.

The constraints which define $\mathcal{E}^*$ are just simple row and column margins. Thus the I-projection, $\mathbb{P}_{\mathcal{E}^*}(g(w))$, can be calculated by the usual IPFP (i.e., adjusting row and column margins), or, as it is a 2×2 table, by direct calculation. As the logarithms of the starting values, w , do not necessarily satisfy the model, the IPFP will in general require several iterations to converge. Thus to obtain the I-projection, $\mathbb{P}_{\mathcal{E}_R}(q_n)$, where $\mathcal{E}_R$ is the space of P.D.'s which satisfy all of the row constraints, we could transform each row of the 3×3 table into a 2×2 table, calculate with the 2×2 table and then use g^{-1} to return a triple of fitted values. The approach for the columns would be similar.

There is another g , which transforms the entire 3×3 table into a $2 \times 2 \times 2 \times 2$ table. In this case $\mathcal{E}^* = g(\mathcal{E})$ becomes the model of no fourth order interaction for the 2^4 table. Specifically,

$$
g : \begin{array}{|c|c|c|} \hline a & b & c \\ \hline d & e & f \\ \hline g & h & i \\ \hline \end{array}
\;\rightarrow\;
\begin{array}{cc}
\begin{array}{|c|c|} \hline a & \frac{1}{2}\,b \\ \hline \frac{1}{2}\,d & \frac{1}{4}\,e \\ \hline \end{array}
&
\begin{array}{|c|c|} \hline \frac{1}{2}\,b & c \\ \hline \frac{1}{4}\,e & \frac{1}{2}\,f \\ \hline \end{array}
\\[1em]
\begin{array}{|c|c|} \hline \frac{1}{2}\,d & \frac{1}{4}\,e \\ \hline g & \frac{1}{2}\,h \\ \hline \end{array}
&
\begin{array}{|c|c|} \hline \frac{1}{4}\,e & \frac{1}{2}\,f \\ \hline \frac{1}{2}\,h & i \\ \hline \end{array}
\end{array}
$$

It is not difficult to check that the model of no fourth order interaction corresponds to $g(E)$ and that $I(p||q) = I(g(p)||g(q))$. Therefore the usual IPFP, with starting values $g(e)$ and the model of no fourth order interaction applied to $g(q_n)$ will yield a 2^4 table of fitted values which can in turn be transformed (by g^{-1}) into a 3 x 3 table for the original problem. ■

Example 4. Paired Comparison Models.

Davidson and Beaver (1977) have considered a generalization of the Bradley-Terry model for paired comparisons which allows for ties and order effects. Fienberg (1979) demonstrated that the models of Davidson and Beaver were loglinear models and showed how the generalized iterative scaling method of Darroch and Ratcliff (1972) can be used for these models. We show how the simple IPFP can also be used to do the estimation.

Consider the $K \times K \times 3$ contingency table $z = \{z_{ijk}\}$ with mean, $m = \{m_{ijk}\}$. The loglinear model corresponding to the Davidson-Beaver model is (see Fienberg (1979)),

$$\ln(m_{ij1}) = \mu + \alpha_{ij} + \beta_1 + \delta_i \ ,$$

$$\ln(m_{ij2}) = \mu + \alpha_{ij} + \beta_2 + \delta_i \ ,$$

and

$$\ln(m_{ij3}) = \mu + \alpha_{ij} + \beta_3 + \tfrac{1}{2}(\delta_i + \delta_j) \ ,$$

for which the sufficient statistics are

$$\{z_{ij+}\} \ , \ \{z_{++k}\} \ , \text{ and } \{z_{i+1} + z_{+i2} + \tfrac{1}{2}(z_{i+3} + z_{+i3})\} \ .$$

Thus the likelihood equations are

$$(4.2) \qquad \hat{m}_{ij+} = z_{ij+} \qquad ij = 1,2,\ldots, K$$

$$(4.3) \qquad \hat{m}_{++k} = z_{++k} \qquad k = 1,2,3$$

and

$$(4.4) \qquad \hat{m}_{i+1} + \hat{m}_{+i2} + \frac{1}{2}(\hat{m}_{i+3} + \hat{m}_{+i3})$$

$$= z_{i+1} + z_{+i2} + \frac{1}{2}(z_{i+3} + z_{+i3}) \qquad i = 1,2,\ldots, K \quad .$$

Fienberg (1979, p. 481) writes out the Darroch and Ratcliff algorithm for this problem.

We transform z into the $K \times K \times 4$ table z^* where

$$(4.5) \qquad z^*_{ij1} = 2 \times z_{ij1}$$

$$(4.6) \qquad z^*_{ij2} = 2 \times z_{ij2}$$

$$(4.7) \qquad z^*_{ij3} = z_{ij3}$$

$$(4.8) \qquad z^*_{ij4} = z_{ij3} \quad , \qquad i,j = 1,2,\ldots, K \quad ,$$

with transformed likelihood equations

$$(4.9) \qquad \hat{m}^*_{ij+} = z^*_{ij+} \qquad i,j = 1,2,\ldots, K$$

$$(4.10) \qquad \hat{m}^*_{++k} = z_{++k} \qquad k = 1,2,3$$

$$(4.11) \qquad \hat{m}^*_{i+1} + \hat{m}^*_{+i2} + \hat{m}^*_{i+3} + \hat{m}^*_{+i3} = z^*_{i+1} + z^*_{+i2} + z^*_{i+3} + z^*_{+i4}$$

$$i = 1,2,\ldots, K$$

$$(4.12) \qquad \hat{m}^*_{i+1} + \hat{m}^*_{+i2} + \hat{m}^*_{i+4} + \hat{m}^*_{+i4} = z^*_{i+1} + z^*_{+i2} + z^*_{i+4} + z^*_{+i4}$$

$$i = 1,2,\ldots, K$$

As the likelihood equations involve simple sums of cell counts, the basic IPFP may be used for this problem. To invert the transformation (4.5) - (4.8) it is necessary that $\hat{m}^*_{ij4} = \hat{m}^*_{ij3}$. Equations (4.9) and (4.12) ensure this. Thus the M.L.E. $\hat{m}$ is

$$(4.13) \qquad \hat{m}_{ij1} = \frac{1}{2}\hat{m}^*_{ij1}$$

$$(4.14) \qquad \hat{m}_{ij2} = \frac{1}{2}\hat{m}^*_{ij2}$$

$$(4.15) \qquad \hat{m}_{ij3} = \hat{m}^*_{ij3} = \hat{m}^*_{ij4}$$

To make the argument rigorous it is necessary to show that if $\hat{m}^*_{ijk}$ satisfy (4.9) - (4.12) then

(i) $\hat{m}^*_{ij3} = \hat{m}^*_{ij4}$

and

(ii) $\hat{m}_{ijk}$ defined by (3.13) - (3.15) satisfy (3.2) - (3.4) .

Condition (i) has already been mentioned and condition (ii) is easily verified by substitution.

This example has again been a case where the transformed table and model are in one to one correspondence with the original table and model. The transformed model can be fitted using the simple IPFP but as the sufficient statistics are not only margins of z^*, many standard computer packages would have difficulty with this problem. ■

5. Social Networks

In recent years there has been an increasing interest in models for the analysis of data from social networks. A line of research described by Holland and Leinhardt (1981) and further developed by Fienberg and Wasserman (1981) and Fienberg, Meyer and Wasserman (1981) has been particularly fruitful.

The basic data for these models consists of observations on the arcs of a directed graph (digraph) on g nodes. The nodes, often taken to represent individuals or organizations in a community, are called actors. The directed arcs linking the actors represent such notions as the attitudes of an individual toward another or the flows of resources between organizations.

A social network with a single relationship connecting actors can be described by an adjacency matrix,

$$X_{ij} = \begin{cases} 1 & \text{if actor } i \text{ connects to actor } j \ (i \rightarrow j) \\ 0 & \text{otherwise} \end{cases}$$

Holland and Leinhardt (1981) develop a model, which they refer to as p_1, and several submodels for such digraph data. Fienberg and Wasserman (1981) extend these models to the case where the actors form disjoint groups and interest lies in the flows between groups. Fienberg, Meyer and Wasserman (1981) further extend these results to the situation where more than one relationship is observed between the actors or groups.

From a computational point of view all of these models are similar. For each of them the likelihood function can be viewed as the Poisson likelihood and the models are either loglinear models or affine transformations of loglinear models for the mean-value parameter. There is a further similarity in that for each case a natural presentation of the data involves non-rectangular data arrays but there exist transformations of the data into rectangular structures for which the transformed sufficient statistics are simple margins. We will consider the simple version of the problem, involving a single relationship between actors and the most general version, involving multiple relations between groups of actors. For these cases we will prove that the simple IPFP can be applied to the transformed data in order to fit the desired models using the method of maximum likelihood.

In order to develop these results we need to consider the original data and distributions. Our presentation will emphasize the mathematical structure, ignoring the interpretation of, and motivation for, the models. We turn first to a development of the Holland and Leinhardt p_1 distribution.

We consider the matrix $X = \{X_{ij}, i = j = 1,2,\ldots, g\}$ as a random matrix to which the distribution will apply. Consider the dyads, or subgraphs, D_{ij} , between actors i and j , where

$$D_{ij} = (X_{ij} , X_{ji}) .$$

The random variable D_{ij} has 4 possible values,

$$D_{ij} = (1,1) : \text{Mutual}$$
$$D_{ij} = (1,0) \text{ or } (0,1) : \text{Asymmetry}$$
$$D_{ij} = (0,0) : \text{Null}$$

Under the assumption of dyadic independence, Holland and Leinhardt (1981) propose the use of the exponential family of distributions,

$$P(X = x) = \exp\left\{ \rho \sum_{i<j} x_{ij} x_{ji} + \theta x_{++} + \sum_i \alpha_i x_{i+} + \sum_j \beta_j x_{+j} \right\} \times K(\rho,\theta,\{\alpha_i\},\{\beta_j\}) \ .$$

Now consider the random variable Y, equivalent to X, which is defined, for $i < j$, as

$$\begin{aligned}
Y_{ij11} &= X_{ij} \cdot X_{ji} \ : \text{Mutual} \\
Y_{ij10} &= X_{ij} \cdot (1 - X_{ji}) \ : \text{Asymmetric} \\
Y_{ij01} &= (1 - X_{ij}) \cdot X_{ji} \ : \text{Asymmetric} \\
Y_{ij00} &= (1 - X_{ij})(1 - X_{ji}) \ : \text{Null}
\end{aligned}$$

corresponding to the values of D_{ij}. Fienberg and Wasserman (1981) show that in terms of Y, the log likelihood function for the model p_1 is:

$$\begin{aligned}
&\ell(\rho,\theta,\{\alpha_i\},\{\beta_j\}|y) \\
&\quad \propto \rho \sum_j \sum_{i<j} y_{ij11} + \theta \sum_j \sum_{i<j} (y_{ij10} + y_{ij01} + 2y_{ij11}) \\
&\quad + \sum_i \alpha_i \left[\sum_{j>i} (y_{ij10} + y_{ij11}) + \sum_{h<i} (y_{hi01} + y_{hi11}) \right] \\
&\quad + \sum_j \beta_j \left[\sum_{i<j} (y_{ij10} + y_{ij11}) + \sum_{j<h} (y_{jh01} + y_{jh11}) \right] .
\end{aligned}$$

Now view y as an element of $\mathcal{Y}$ = {p.m.f.'s on the index set $\mathcal{K}$} where $\mathcal{K} = \{(i,j,k,\ell);\ i < j = 1,2,\ldots, g;\ k,\ell = 0,1\}$. If we consider y to be distributed as a collection of independent Poisson random variables with mean $q \in \mathcal{Y}$ then the likelihood is exactly that which would be obtained by using the loglinear model

$$\ln(q) \in \mathcal{M} \subset R^{\mathcal{K}}.$$

The manifold, $\mathcal{M}$, is spanned by the vectors $f^{\delta} \in R^{\mathcal{K}}$, $\delta = 1,2,\ldots, 2 + 2g$, given by

$$(5.1) \qquad \rho \qquad f^1 = \begin{cases} 1 : (k,\ell) = (1,1) \\ 0 : \text{otherwise} \end{cases}$$

$$(5.2) \qquad \theta \qquad f^2 = \begin{cases} 2 : (k,\ell) = (1,1) \\ 1 : (k,\ell) = (1,0) \text{ or } (0,1) \\ 0 : \text{otherwise} \end{cases}$$

$$(5.3) \qquad \alpha_{i'} \quad f^{2+i'} = \begin{cases} 1 : (k,\ell) = (1,1) \text{ or } (k,\ell) = (1,0) \text{ and} \\ \quad\ j > i',\ i = i' \text{ for } (k,\ell) = (1,0) \text{ and} \\ \quad\ j = i',\ i < j \\ 0 : \text{otherwise} \end{cases} \qquad i' = 1,2,\ldots, g.$$

$$(5.4) \qquad \beta_{j'} \quad f^{2+g+j'} = \begin{cases} 1 : (k,\ell) = 1,1) \text{ or } (k,\ell) = (1,0) \text{ and } j = j', \\ \quad\ i > j' \text{ or } (k,\ell) = (0,1) \text{ and } i = j',\ j < i \\ 0 : \text{otherwise} \end{cases} \qquad j' = 1,2,\ldots, g$$

This spanning set was chosen so that the inner product of an observed y with the f's yields the sufficient statistics:

$$(5.5) \qquad \rho \qquad a^1 = \sum_j \sum_{i<j} y_{ij11}$$

$$(5.6)\quad \theta \qquad a^2 = \sum_j \sum_{i<j} (y_{ij10} + y_{ij01} + 2y_{ij11})$$

$$(5.7)\quad \alpha_{i'} \qquad a^{2+i'} = \sum_{j>i'} (y_{i'j10} + y_{i'j01}) + \sum_{h<i'} (y_{hi'01} + y_{hi'11})$$
$$i' = 1,2,.., g$$

$$(5.8)\quad \beta_{j'} \qquad a^{2+g+j'} = \sum_{i<j'} (y_{ij'10} + y_{ij'11}) + \sum_{h>j'} (y_{j'h01} + y_{j'h11})$$
$$j' = 1,2,.., g$$

We now collect the spanning vectors into $F = \{f^h : h = 1,2,\ldots,(2+2g)\}$ and the observed sufficient statistics into $A = \{a^h : h = 1,2,\ldots,(2+2g)\}$. If we define the linear space of P.D.'s, E, by the constraints, F, and corresponding constants, A, then the M.L.E. is

$$\hat{q} = \mathbb{P}_E(r)$$

where $r \in Y$ and $r_{ijk\ell} = c \;\; \forall (ijk\ell) \in K$. Thus a natural setting for the estimation of p_1 is as a loglinear model on Y. As the vector f^2 is not a zero-one vector, and cannot be cast in this form, the basic IPFP can not be used for the estimation problem. In addition for many problems g will be so large that Newton's method can not be used. It would be desirable if the problem could be put in a form where a standard algorithm could be used.

The space Y is a rather convoluted construction. It would be more natural to work with $Y^* = \{$p.m.f.'s on the index set $K^*\}$ where $K^* = \{(i,j,k,\ell) : i,j = 1,2,\ldots, g ; k,\ell = 0,1\}$, the space of $g \times g \times 2 \times 2$ tables. To this end consider the transformation

$g : \mathcal{Y} \to \mathcal{Y}^*$ with $y^* = g(y)$ defined by

$$y^*_{ji\ell k} = \begin{cases} y_{ijk\ell} & i < j \\ 0 & i = j \end{cases}$$

In other words we have transformed the problem into a $g \times g \times 2 \times 2$ contingency table with zeros on the "diagonals". The sufficient statistics (5.5) - (5.8) appear (sometimes more than once) as the [12], [13], [14], [23], [24], and [34] margins of y^*. Now consider the linear space of p.m.f.'s, $\mathcal{E}^*$, defined by

$$\mathcal{F}^* = \{[12], [13], [14], [23], [24], \text{ and } [34] \text{ margin functions}\}$$

and

$$\underline{A}^* = \{[12], [13], [14], [23], [24], \text{ and } [34] \text{ margins of } y^*\} .$$

We should note that $\mathcal{E}^*$ is not equal to $g(\mathcal{E})$. In fact,

$$g(\mathcal{E}) = \mathcal{E}^* \cap \{y^*_{ijk\ell} : y^*_{ijk\ell} = y^*_{ji\ell k}\} .$$

In other words $g(\mathcal{E})$ is a strict subset of $\mathcal{E}^*$. As the model, $\mathcal{E}^*$, requires just simple margins of a rectangular data array, the basic IPFP found in many computer packages can be used. We would like to be able to fit just $\mathcal{E}^*$ to y^*, ignoring the symmetry constraints. Let

$$\hat{q}^* = \mathbb{P}_{\mathcal{E}^*}(g(r))$$

where

$$r^*_{ijk\ell} = g(r) = \begin{cases} c & i \neq j \\ 0 & i = j \end{cases}$$

As $\hat{q}^*$ is easy to calculate we would like to assert that $\hat{q}^* \in g(\mathcal{E})$.

One method of proceeding would be to go ahead and fit $\mathcal{E}^*$ to y^*. If $\hat{q}^*$ has the desired symmetries then all is well. In general we need to prove that for an arbitrary y, $\hat{q}^*$ must be in $g(\mathcal{E})$.

Our first version of this proof relied upon the actual calculations involved in the IPFP to show the symmetry. The proof presented here is much simpler and relies only on an invariance argument.

Let h denote the mapping from $R^{g\times g\times 2\times 2}$ into $R^{g\times g\times 2\times 2}$ defined by

$$h : z_{ijk\ell} \to z_{ji\ell k} ,$$

i.e., the symmetry transformation. In order that $\hat{q}^*$ be in $g(\mathcal{E})$ we require that

$$h(\mathbb{P}_{\mathcal{E}^*}(g(r))) = \mathbb{P}_{\mathcal{E}^*}(g(r)) .$$

Now notice that

$$h([12] \text{ margin function}) = [12] \text{ margin function} ,$$

$$h([13] \text{ margin function}) = [24] \text{ margin function} ,$$

and that each of the other margin functions in F^* is mapped into another margin function in F^*. Similarly

$$h([13] \text{ margin for data } y^*) = [24] \text{ margin for data } y^* .$$

In other words, $h(F^*) = F^*$ and $h(A^*) = A^*$ which together imply that $h(\mathcal{E}^*) = \mathcal{E}^*$. Also note that $h(g(r)) = g(r)$. We can then assert that

$$\hat{q}^* = \mathbb{P}_{\mathcal{E}^*}(g(r)) = \mathbb{P}_{h(\mathcal{E}^*)}(h(g(r))) = h(\hat{q}^*)$$

and hence the result.

We have now shown that the M.L.E. $\hat{q}^*$ resulting from fitting $\mathcal{E}^*$ to y^* is in $g(\mathcal{E})$ and hence $\hat{q} = g^{-1}(\hat{q}^*)$.

There are numerous submodels of p_1 considered by Holland and Leinhardt (1981) and Fienberg and Wasserman (1981). These models, represented in terms of parameters and margins in the y^* table are listed in Table 5.1.

Table 5.1 Submodels of p_1

Special Case	Parameters	Margins Fitted
(i)	$\rho, \theta, \{\alpha_i\}, \{\beta_j\}$	[12] [13] [14] [23] [24] [34]
(ii)	$\theta, \{\alpha_i\}, \{\beta_j\}$	[12] [13] [14] [23] [24]
(iii)	$\rho, \theta, \{\alpha_i\}$	[12] [13] [24] [34]
(iv)	$\theta, \{\alpha_i\}$	[12] [13] [24]
(v)	$\rho, \theta, \{\beta_j\}$	[12] [14] [23] [34]
(vi)	$\theta, \{\beta_j\}$	[12] [14] [23]
(vii)	ρ, θ	[12] [34]
(viii)	θ	[12] [3] [4]

Each of these sets of margins are invariant under h and the above argument is applicable.

For the p_1 problem all of the models in Table 4.1 can be fit using the basic IPFP on the data y^*.

Our second example concerns a class of loglinear models for multivariate directed graphs as described in Fienberg, Meyer and Wasserman (1981). They consider a set of data concerning the inter-relationships between 73 organizations in a small community. Three types of relationships were observed for each of the pairs of organizations, but for simplicity we restrict our attention to two of these criteria, support and money. For each criterion the organizations were asked to respond to the questions:

(i) to which organizations do you give support (money)?

(ii) from which organizations do you receive support (money)?

A particular directed relationship (i.e., giving or receiving) is regarded to be present if either or both the organizations in a pair perceived the relationship. For each pair of organizations it is possible to construct a four-vector of zeros and ones indicating the presence or absence of (support out, support in, money out, money in). Consider for the moment just the support relationship. A pair of organizations are said to have a Mutual relationship if they support each other (i.e., (support out, support in) = (1,1)), a Null relationship if neither supports the other (i.e., (0,0)), or an Asymmetric relationship if support is unreciprocated (i.e., (0,1) or (1,0)). If we aggregate over all $\binom{73}{2} = 2628$ pairs of organizations there are ten distinguishable support-money relationships, namely,

MM	with four vector	(1,1,1,1)
MA		(1,1,0,1) or (1,1,1,0)
MN		(1,1,0,0)
AM		(0,1,1,1) or (1,0,1,1)
AA		(0,1,0,1) or (1,0,1,0)
$A\bar{A}$		(0,1,1,0) or (1,0,0,1)
AN		(0,1,0,0) or (1,0,0,0)
NM		(0,0,1,1)
NA		(0,0,1,0) or (0,0,0,1)
NN		(0,0,0,0)

Notice that when both relationships are asymmetric there are two different cases, corresponding to whether the relationships flow in the same or in different ways. We denote the table of observed probabilities by z where for example z_{MM} is the number of mutual-mutual relationships divided by $\binom{73}{2}$. The table is represented by

$z =$

<table>
<tr><td></td><td></td><td colspan="3">MONEY</td></tr>
<tr><td></td><td></td><td>M</td><td>A</td><td>N</td></tr>
<tr><td rowspan="4">SUPPORT</td><td>M</td><td>z_{MM}</td><td>z_{MA}</td><td>z_{MN}</td></tr>
<tr><td rowspan="2">A</td><td rowspan="2">z_{AM}</td><td>z_{AA}</td><td rowspan="2">z_{AN}</td></tr>
<tr><td>$z_{A\bar{A}}$</td></tr>
<tr><td>N</td><td>z_{NM}</td><td>z_{NA}</td><td>z_{NN}</td></tr>
</table>

Fienberg, Meyer and Wasserman (1981) model the probability, $q = \{q_{ab}\ ;\ a,b = M,A,N\}$ that a randomly selected dyad will be assigned to a certain cell. They consider linear models for $\xi = \{\xi_{ab}\ ;\ a,b = M,A,N\}$ where

$$\xi_{ab} = \begin{cases} \log(q_{ab}) & \text{if } a,b \text{ each equal } M \text{ or } N \\ \log(q_{ab}/2) & \text{if either } a \text{ or } b \text{ equals } A\ . \end{cases}$$

These models are affine translations of loglinear models for q. The arguments presented here apply to all of their models.

The model we consider takes as a linear space, E, of p.m.f.'s the set of tables, s, which have margins s_{a+} and s_{+b}, $a,b = M,A,N$, which are the same as the corresponding margins for the z-table. For example we require

$$s_{A+} = s_{AM} + s_{AA} + s_{A\bar{A}} + s_{AN} = z_{AM} + z_{AA} + z_{A\bar{A}} + z_{AN} = z_{A+}\ .$$

In order to have the model be linear in ξ, we need

$$\hat{q} = \mathbb{P}_E(r)$$

where

$$r_{ab} = \begin{cases} 1 & \text{if } a,b \text{ each equal M or N} \\ \frac{1}{2} & \text{if either a or b equal A}\ . \end{cases}$$

As the model space can be spanned by vectors consisting of 0's and 1's, the simple IPFP, which takes an initial table, r, and successively adjusts the row and column "margins" to match those in the observed table, can be used. This algorithm is easy to do by hand, but because the z-table is not rectangular (i.e., it has 10 cells rather than the 9 one would expect), and consequently has an extended interpretation of margin totals, many standard IPFP computer programs would not be able to analyze this table. Moreover, for many of the models considered by Fienberg, Meyer and Wasserman the models are not so simple and the computations on the z-table require more than the simple IPFP . For this reason we prefer to work with a transformed problem, where the sufficient statistics for the models can be represented by simple marginal totals.

An alternate, though somewhat deceptive, description of the data is to consider four-vectors for each of the $\binom{73}{2} \times 2$ ordered pairs of organizations and to aggregate this into a 2^4 table, $y = y_{ijk\ell}$, $i,j,k,\ell = 1,2$, where a 1 indicates the presence of a flow and a 2 indicates the absence of a flow. Thus y_{1111} is the number of mutual-mutual relationships divided by 5256. The y table duplicates certain relationships and gives double weight to certain others. The y-table has the form,

	money out	1		2	
	money in	1	2	1	2
supp out	supp in				
1	1	y_{1111}	y_{1112}	y_{1121}	y_{1122}
	2	.	.	.	.
2	1	.	.	.	.
	2	.	.	.	.

We now consider the transformation which maps the z-table into the y-table; viz.,

$$
g : z \to \frac{1}{2}
\begin{array}{|c|c||c|c|}
\hline
2z_{MM} & z_{MA} & z_{MA} & 2z_{MN} \\
\hline
z_{AM} & z_{AA} & z_{A\bar{A}} & z_{AN} \\
\hline
\hline
z_{AM} & z_{A\bar{A}} & z_{AA} & z_{AN} \\
\hline
2z_{NM} & z_{NA} & z_{NA} & 2z_{NN} \\
\hline
\end{array}
$$

$$
= \frac{1}{2} y \quad .
$$

We denote the factors support (out, in), money (out, in) by the numbers 1, 2, 3, and 4. It is now easy to see that the marginal sums considered for the z-table can all be found (twice) in the [12] and [34] margins of the y-table. Also note that the y-table has a strong symmetry, $y_{ijk\ell} = y_{ji\ell k} \; \forall \, ijk\ell$. Now $g(\mathcal{E})$ is just the set of tables which have (i) the correct [12] and [34] margins and (ii) preserve the observed symmetry in the y-table. Consider just the first of these conditions ignoring the symmetry constraint. It is this model which we shall consider to be $\mathcal{E}^*$. As we have relaxed some conditions it is clear that $g(\mathcal{E}) \subset \mathcal{E}^*$.

From here on the argument proceeds in the same manner as in the single relationship case. It is convenient now to explicitly define the space $\mathcal{E}^*$ and the conditions we need to verify to show that $\mathbb{P}_{\mathcal{E}^*}(g(r))$ is in $g(\mathcal{E})$. Consider

$$
F^* = \{f_1, \ldots, f_8\} \text{ where}
$$

$$f_1 = \begin{array}{|c|c|c|c|}\hline 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline \end{array} \quad \cdots \quad f_4 = \begin{array}{|c|c|c|c|}\hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 1 & 1 & 1 & 1 \\ \hline \end{array}$$

$$f_5 = \begin{array}{|c|c|c|c|}\hline 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ \hline \end{array} \quad \cdots \quad f_8 = \begin{array}{|c|c|c|c|}\hline 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array}$$

and constants $\mathcal{A}^* = \{a_1, \ldots, a_8\}$ where $a_j = \langle f_j, g(z)\rangle$. Note that $a_2 = a_3$ and $a_6 = a_7$. We define $\mathcal{E}^*$ to be the space of P.D.'s defined by $\mathcal{F}^*$ and $\mathcal{A}^*$. Now consider the symmetry transformation:

$$h : y_{ijk\ell} \to y_{ji\ell k} .$$

For $\mathbb{P}_{\mathcal{E}*}(g(r))$ to be in $g(\mathcal{E})$ we require

$$h(\mathbb{P}_{\mathcal{E}*}(g(r)) = \mathbb{P}_{\mathcal{E}*}(g(r)) .$$

It is possible to assert this because the space $\mathcal{E}^*$ is invariant under h. Specifically $h(f_i) = f_i$ for i = 1,4,5,8 and $h(f_2) = f_3$, $h(f_3) = f_2$, $h(f_7) = f_6$ and $h(f_6) = f_7$. Because $a_2 = a_3$ and $a_6 = a_7$ the linear space $h(\mathcal{E}^*)$ generated by $h(F^*)$ and $h(\mathcal{A}^*)$ is the same as $\mathcal{E}^*$. We also note that $h(g(r)) = g(r)$, because of the nature of g function. That is the starting values necessarily satisfy the symmetry constraints. Now let

$$\hat{q}^* = \mathbb{P}_{\mathcal{E}^*}(g(r)) \quad \text{and}$$

$$\tilde{q}^* = \mathbb{P}_{h(\mathcal{E}^*)}(h(g(r))) = \mathbb{P}_{\mathcal{E}^*}(g(r)) \; .$$

But note that $\hat{q}^* = h(\hat{q}^*)$ as all we have done is relabel the co-ordinates. Thus

$$\hat{q}^* = \tilde{q}^* = h(\hat{q}^*)$$

i.e., the fitted P.D. is (i) invariant under h and (ii) is in $\mathcal{E}^*$. Thus $\hat{q}^*$ is in $g(\mathcal{E})$ and $g^{-1}(\hat{q}^*)$ is the fitted P.D. in the space of Z-tables.

For any of the other models considered by Fienberg, Meyer and Wasserman, it is easy to show that the space, $\mathcal{E}^*$, is invariant under h and thus the above argument still works.

In these examples, $g(r)$ is the uniform distribution; thus the IPFP with starting value all ones is an appropriate algorithm. For some of the models, the appropriate margins of the y*-table represent a decomposable model; in fact the model [12], [34] is itself decomposable. Thus we have not only found an easy computational procedure, but have also discovered closed-form estimates for some of the models. The existence and nature of closed-form estimates varies with the number of relationships between actors which are modeled.

The analysis of the multiple relationship data that we have considered has been for the data aggregated over all the actors. In some situations it may be desirable to aggregate over only groups of actors, in which case there is a 2^4 (or with 3 relationships, 2^6) table for each group of actors. In this manner it is possib'e for the number of entries in the table, and the number of parameters in the corresponding models, to grow very large. Under these circumstances the transformation techniques outlines in this chapter prove to be of considerable practical use.

6. Desiderata

We conclude this chapter with a few questions and cautions. The examples have shown situations where, for reasons of computational ease, it was desirable to transform a contingency table into a related but larger table. In the transformed table it was possible to fit a model using the standard IPFP whereas in the original table the corresponding model would have required a more complicated algorithm. This approach of using transformed tables is especially important in practice as versions of the standard IPFP are widely available and easy to use. An additional bonus which can sometimes be found in the transformed table is the existence of closed form maximum likelihood estimates. The theory about when closed form estimates exist in complete tables with factorial models is well known and such situations are easily recognized. On the contrary, when a table is incomplete or has a more complicated structure, very little is known about the existence of closed form estimates. Our techniques have merely scratched the surface of the more general question of closed form estimates. A more general theory of closed form estimates for arbitrary loglinear models would seem desirable; perhaps investigations of the more general IPFP will aid in this.

Throughout our discussion we have ignored the important questions of degree of freedom calculations and asymptotic covariance estimates for the M.L.E. When $g(E) = E^*$, that is we are essentially only relabeling the problem, then any d.f. and covariances calculated in E^* can be transformed back to E. When $g(E) \subset E^*$, special care must be taken to calculate the appropriate d.f. in E. We know of no exact procedure for transforming covariance estimates in E^* back to E and suspect that it is not possible.

Acknowledgments

This research was partially supported by the Office of Naval Research Contract N00014-80-C-0636 at Carnegie-Mellon University. I would like to thank S.E. Fienberg and S.S. Wasserman for helpful discussions during the preparation of this work.

References

Csiszár, I. (1975), I-Divergence Geometry of Probability Distributions and Minimization Problems. The Annals of Probability, 3:146-158.

Darroch, J. N. and Ratcliff, D. (1972), Generalized Iterative Scaling for Log-linear Models. The Annals of Mathematical Statistics, 43:1470-1480.

Davidson, R. R. and Beaver, R. J. (1977), On extending the Bradley-Terry Model to Incorporate Within-pair Order Effects. Biometrics, 33:693-702.

Fienberg, S. E. (1979), Loglinear Representation for Paired Comparison Models with Ties and Within-pain Order Effects. Biometrics, 35:479-481.

Fienberg, S. E., Meyer, M. M., and Wasserman, S. S. (1981), Analyzing Data from Multivariate Directed Graphs: An Application to Social Networks. To appear in Vic Barnett (ed.), Looking at Multivariate Data. Chichester, England: John Wiley & Sons.

Fienberg, S. E. and Wasserman, S. S. (1981), Categorical Data Analysis of Single Sociometric Relationships. Sociological Methodology 1981. San Francisco: Jossey-Bass.

Gokhale, D. V. and Kullback, S. (1978), The Information in Contingency Tables. New York: Marcel Dekker.

Haberman, S. J. (1974), The Analysis of Frequency Data. Chicago: The University of Chicago Press.

Holland, P. W. and Leinhardt, S. (1981), An Exponential Family of Probability Distributions for Directed Graphs. Journal of the American Statistical Association, 76:33-50.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 214	2. GOVT ACCESSION NO. AD-A104 282	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 TRANSFORMING CONTINGENCY TABLES		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER 9 Technical Report, No. 214
7. AUTHOR(s) 10 Michael M. Meyer		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-80-C-0636
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Carnegie-Mellon University Pittsburgh, PA 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 12 39
11. CONTROLLING OFFICE NAME AND ADDRESS Contracts Office Carnegie-Mellon University Pittsburgh, PA 15213		12. REPORT DATE 11 August, 1981
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 14 TR-214		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Iterative Proportional Fitting, Kullback-Leibler Information Number, Social Networks

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We consider methods for transforming contingency table and associated models into a form where the models and computations are simpler. This technique can often help in the recognition of closed-form estimates. Major examples are drawn from the theory of social network modelling.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

391190

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DATE
ILMED
-8